



## Machine Readable Archives — Two Steps

When I originally proposed this article, I had in mind a discussion about the declining cost of running a machine readable archives program. The Government Archives Division, along with other divisions at the National Archives, has overhauled the way that it deals with "computer records," increasing capability while decreasing cost. As I thought about it, though, I concluded that what I really want to do is convince you, the audience, that NOW is an excellent time to start such a program.

Basically, there are two steps involved in setting up a machine readable archives: **get a computer; then get some data.**

### Get a Computer

Until about two years ago, if you wanted to start a machine readable archives, it was almost essential to have access to a mainframe, or at least a minicomputer. This was, and still is, expensive — for computer and software purchases and for training. However, over the last few years, so-called microcomputers have become powerful enough to be practical for machine readable records processing. And compared to minis and mainframes they are cheaper. As an example, the National Archives spent more than \$250,000 in 1989 for the rental of mainframe computer facilities. The cost to purchase the microcomputer system now in use was less than \$120,000. The system is expected to last at least three years. So, a conservative annual cost would be \$45,000, taking into account normal operating expenses. The benefits of the change go well beyond simple dollar savings. There are programs available on microcomputers that do not exist on minis and mainframes. Even when they do exist, they are ten to two hundred times more expensive on minis and mainframes than on micros. So it is much cheaper to provide access to records through microcomputers.

### Get Some Data

One of the problems facing an archivist when assessing a database (a collection of structured electronic records) is answering the question, "Is there anything valuable in here?" A program called Knowledge Seeker is most helpful in answering that question. It does an analysis of the database automatically and tells you all the statistically significant things it finds out about the database contents. You do not have to be a statistician to use it. You do have to decide if there is any archival value in

what it discovers. The program costs about \$600.

Another problem that has been discussed at great length is the lack of standards for electronic text. The chaos of formats, the lack of organization of electronic documents, the difficulty of separating "routine" from "important" have all been lamented. Meanwhile, most organizations in Canada are using computers to do their "paperwork." Our experience with the records of the Trade Negotiations Office proves that electronic documents can be acquired and made available to researchers. Software such as Magellan (\$180) can cope with the formats. Since it uses full text search (that is, searches the complete contents of all documents), the lack of a file number or other organizer is less of a problem. It also makes it cheaper to take the whole collection rather than to try to separate out and transfer the valuable documents.

A third barrier to machine readable archives has been preservation. The traditional approach to preservation is to use magnetic tape, which requires climate controls, rewinding and recopying. An alternative is the optical disk. Optical disk storage devices are available for microcomputers. The disks themselves store large quantities of data. They need no special environment or handling, and they last up to fifty years (well beyond the life of the machines they are used in). True, this is not as long as paper, but the shorter lifespan is partially offset by lower costs of storage space. The one caveat about optical disk is to make sure that the one you select conforms to an international standard, because there are still many proprietary formats around and you want to avoid these for use in machine readable archives.

### Miscellaneous

Up to this point I have discussed some barriers to setting up a successful machine readable archives, and how the evolution of microcomputer technology has reduced or eliminated them. In any endeavour, the people involved are the critical factor. In reviewing the fifteen-year experience of the National Archives with machine readable records, I conclude that any such operation requires some people with technical expertise. How this expertise is combined in an archives is a matter for debate. The U.S. (National Archives and Records Administration) approach has been to create a special unit with total responsibility for archival and technical judgements. In the Government Archives Division a separate organization was set up to look after technical issues, while the archivists dealt with archival

issues. There are advantages to both approaches, but the separation of the technical from the archival issues makes it easier to establish the necessary technical expertise. Good technical people are in demand these days.

Who do I consider are "good technical people?" I look for three things. I prefer someone who enjoys computer adventure games. The reality of acquiring computer records is this — more than half the data is not what it was supposed to be. Thus much of the technical work involves playing detective to determine what you really have. This turns out to be a lot like solving the puzzles in an adventure game. Next, I look for someone who has experience with a large variety of computer systems. The federal government has bought at least one of every computer ever sold in North America. Knowing IBM or Digital or Sperry inside out is less useful in archives than knowing something about all three (and a bunch more). Finally, I would like some indication that the person is partly self-taught. Any archives has to keep up with its donors' technologies and computer evolution in general. Formal courses are expensive, and I have not seen any that deal effectively with the day to day demands of a machine readable archives. In any case, it is always the people who make a program successful.

### Summing Up

Certainly, there are lots of issues about computer records that can be debated. There are some problems that need to be solved. In the meantime, valuable records are being destroyed by organizations that no longer need them. Take some action to save them: **get a computer; then get some data.**

Information concerning the software to which reference is made in this article can be obtained by contacting the following:

Knowledge Seeker  
Firstmark Technologies  
14 Concourse Gate, Suite 600  
Ottawa, Ontario  
K2E 7S8  
Bus: (613) 723-8020  
Fax: (613) 723-8048

Magellan  
Lotus Corporation  
Available through most retail computer suppliers.

Doug Taylor-Munro  
Government Archives Division



## GIS at the National Archives

Of the 35 federal institutions surveyed last year by the Inter-Agency Committee on Geomatics, the National Archives of Canada placed among the top five in terms of the number of datasets it retains. "And the figure is only going to get bigger," replies David Brown, an archivist with the Cartographic and Audio-Visual Archives Division. "Under the new *Archives Act* passed in 1987, the National Archives of Canada was given a very clear mandate to collect the memory of the government of Canada, no matter what the media."

The *Act* referred to by Brown requires all federal departments to obtain the permission of the National Archivist before destroying their records, whether they take the form of electronic data, paper records, maps, photographs or films. If the records have historical, legal or evidential value, the *Act* also provides for their transfer to the National Archives of Canada for permanent retention once they are no longer considered operational by the department. With the *National Archives of Canada Act* now enshrined in law, the Archives has collected some 13 GIS (Geographic Information System) datasets from a variety of federal institutions, most notably Environment Canada, Forestry Canada, and the Geological Survey of Canada.

To some, this must seem an odd contradiction. The institution that is well-known for its collections of the oldest paper-based records of the government is now turning its attention to the most sophisticated, electronic databases ever created. "Our goal is the retention of Canada's national memory," reiterates senior archivist Terry Cook. "Since Canada is a world leader in the development of GIS, we are quite prepared to take whatever steps are necessary to document this achievement for future generations."

Unfortunately, electronic records are extremely fragile and transient, and many of the original datasets that were created back in the 1970s are now lost forever. This loss is partially the result of the rather callous attitudes many departments had for their information. Senior officials usually had no idea what data were being collected by their department, let alone where they might find it. Since government departments had little interest in managing their information, decisions on the retention of electronic records were generally left to the discretion of junior personnel who had no concept as to how the information might be used by future researchers.

But with the help of Treasury Board's new policy on the Management of Government Information Holdings, federal departments are now beginning to recognize their electronic records as a corporate asset. The new policy calls for a "designated senior official" to represent the department and take responsibility for

implementing a corporate perspective on the management, and later efficient disposition, of all information in all media within his or her organization.

This policy has established new precedents, and it has helped the National Archives of Canada to be one of the first in the world to collect electronic records and GIS datasets. In combination, the *National Archives of Canada Act* and Treasury Board's policy have enabled the National Archives to implement a planned approach to the disposition of government records. Over the next few years, archivists will be reviewing all the holdings of major government departments in order to identify records of national significance. The result will hopefully enable a more complete and systematic preservation of GIS datasets throughout federal institutions.

Further information on the GIS holdings of the National Archives of Canada can be obtained by writing to the Director, Reference and Researcher Services Division, National Archives of Canada, 395 Wellington Street, Ottawa, Ontario, K1A 0N3.

Jeffrey Murray  
Government Archives Division

Anyone wishing information about the *Bulletin* may write to: National Archives of Canada, Sam Kula, Assistant Director General, Historical Resources Branch, 395 Wellington Street, Ottawa, Ontario K1A 0N3, or phone (613) 996-1579.



## En conclusion

Il y a bien sûr beaucoup de questions relatives aux archives informatiques dont on peut discuter. Certains problèmes doivent être résolus. Pendant ce temps, des fichiers importants sont détruits par des organismes qui n'ont plus besoin. Faites quelque chose pour les sauver : procurez-vous un ordinateur, puis obtenez un certain nombre de données.

Vous pouvez obtenir des renseignements supplémentaires au sujet des logiciels mentionnés dans le présent article en communiquant avec les entreprises suivantes :

Knowledge Seeker  
Firstmark Technologies  
14, Concourse Gate, Bureau 600  
Ottawa (Ontario)  
K2E 7S8  
Téléphone : (613) 723-8020  
Télocopieur : (613) 723-8048

Magellan  
Lotus Corporation  
Vous pouvez obtenir ce logiciel dans la plupart des boutiques qui vendent des produits d'informatique.

Doug Taylor-Munro  
Division des archives gouvernementales

## Le SIG aux Archives nationales

Parmi les 35 organismes gouvernementaux sondés l'an dernier par le Comité mixte des archives nationales du Canada figuraient parmi les cinq meilleurs en ce qui a trait au nombre de fichiers conservés. « Et ce nombre ne fera qu'augmenter », affirme David Brown, archiviste de la Division des archives cartographiques et audiovisuelles. « En vertu de la

nouvelle Loi sur les Archives nationales du Canada adoptée en 1987, les AN ont obtenu le mandat très clair de réunir la mémoire du gouvernement du Canada, peu importe le support. »

La Loi mentionnée par M. Brown oblige tous les ministères fédéraux à obtenir l'autorisation de l'Archiviste national avant de détruire leurs documents, qu'il s'agisse de données informatiques, de documents en papier, de cartes géographiques, de photographies ou de films. Si les documents ont une valeur historique, juridique ou probatoire, la Loi prévoit aussi leur transfert aux AN pour conservation permanente, dès que le ministre considère qu'ils ne sont plus des documents courants. Maintenant que la Loi a été adoptée, les Archives ont regroupé 13 fichiers du SIG (Système d'information géographique) provenant de diverses institutions fédérales, notamment Environnement Canada, Forêts Canada et la Commission géologique du Canada.

Aux yeux de certaines personnes, cela peut sembler une contradiction étrange. L'établissement, qui est bien connu pour ses fonds des plus anciens dossiers sur support en papier du gouvernement, dirige maintenant son attention vers les bases de données les plus perfectionnées jamais créées. Selon l'archiviste Terry Cook, notre but est la conservation de la mémoire nationale du Canada. « Étant donné que le Canada est un chef de file mondial dans le développement du SIG, nous sommes plutôt bien préparés pour entreprendre les étapes nécessaires en vue de documenter cette réalisation pour les générations futures. »

Malheureusement, les fichiers informatiques sont extrêmement fragiles et éphémères, et un grand nombre des fichiers originaux, créés au cours des années 1970, sont perdus pour toujours. Cette perte résulte en partie des attitudes plutôt insensibles que les ministères adoptaient à l'égard de leur fonds de renseignements. Les hauts fonctionnaires n'avaient habituellement aucune idée des données qui avaient été recueillies par leur ministère, et encore moins de l'endroit où ils pourraient les trouver. Étant donné que les

ministères montraient peu d'intérêt pour la gestion de leur fonds de renseignements, les décisions relatives à la conservation des fichiers informatiques étaient habituellement laissées à la discrétion du personnel subalterne, qui ne savait absolument pas de quelle façon ces renseignements pourraient être utilisés ultérieurement par des chercheurs.

Avec l'aide de la nouvelle politique du Conseil du Trésor, *La gestion des renseignements détenus par le gouvernement*, les ministères fédéraux commencent maintenant à considérer leurs fichiers informatiques comme un actif. La nouvelle politique exige qu'un « haut fonctionnaire nommé » représente le ministre et prenne la responsabilité de la mise en œuvre d'une perspective institutionnelle pour la gestion, et la disposition future efficiente de tous fonds de renseignements dans tous les médias à l'intérieur de son organisme.

Cette politique a créé de nouveaux précédents et a aidé les AN à être parmi les premiers organismes au monde à conserver des fichiers informatiques et des fichiers du SIG. De plus, la Loi ainsi que la politique du Conseil du Trésor ont permis aux AN de mettre en œuvre une approche structurée pour la disposition des documents gouvernementaux. Au cours des prochaines années, les archivistes examineront toutes les ressources documentaires des principaux ministères afin de repérer les documents d'importance nationale. Il est à souhaiter que le résultat permettra la conservation plus complète et systématique des fichiers du SIG pour l'ensemble des institutions fédérales.

Vous pouvez obtenir des renseignements supplémentaires au sujet des ressources documentaires du SIG aux AN en écrivant au Directeur, Division des services de référence et des services aux chercheurs, Archives nationales du Canada, 395, rue Wellington, Ottawa (Ontario), K1A 0N3.

Jeffrey Murray  
Division des archives gouvernementales





# Archives informatiques



## Archives informatiques — Deux étapes

Lorsque j'ai d'abord proposé cet article, j'avais l'idée de traiter du coût d'exploitation décroissant d'un programme d'archives informatiques. La Division des archives gouvernementales, de concert avec d'autres divisions des Archives nationales (AN), a révisé la façon dont elle s'occupe des « dossiers informatiques », en augmentant la productivité tout en diminuant les coûts. En y repensant bien, par contre, j'estime que mon but véritable est de vous convaincre que MAINTENANT est un excellent moment pour lancer un tel programme.

Fondamentalement, la mise en place d'archives informatiques comprend deux étapes : se procurer un ordinateur; puis obtenir un certain nombre de données.

### Se procurer un ordinateur

Jusqu'à il y a environ deux ans, si vous désirez créer des archives informatiques, vous deviez presque absolument avoir accès à un ordinateur central, ou au moins à un mini-ordinateur. Les coûts relatifs à l'achat de logiciels et d'ordinateurs, ainsi qu'à la formation et d'ordinateurs, étaient, et sont toujours, élevés. Cependant, au cours des dernières années, les micro-ordinateurs sont devenus assez puissants pour servir au traitement des archives informatiques. De plus, comparativement aux mini-ordinateurs et aux ordinateurs centraux, ils sont moins coûteux. Par exemple, les AN ont dépensé plus de 250 000 \$ en 1989 pour la location d'installations informatiques. Le coût d'achat des micro-ordinateurs utilisés à l'heure actuelle était inférieur à 120 000 \$. On prévoit qu'ils dureront au moins trois ans. Par conséquent, le coût annuel approximatif serait de 45 000 \$, en tenant compte des frais d'exploitation habituels. Les avantages liés au changement vont beaucoup plus loin que les simples économies en argent. Certains programmes pour micro-ordinateurs n'existent pas pour les mini-ordinateurs et les ordinateurs centraux. Même lorsqu'ils existent, ils coûtent de dix à deux cents fois plus cher pour les mini-ordinateurs et les ordinateurs centraux que pour les micro-ordinateurs. Il est donc moins coûteux d'utiliser les micro-ordinateurs afin de permettre l'accès aux archives.

### Obtenir un certain nombre de données

Trouver la réponse à la question : « Y a-t-il quelque chose de valeur pour la recherche dans la base ? » est l'un des problèmes aux-

quels fait face l'archiviste dans l'évaluation

d'une base de données (une collection de fichiers informatiques structurés). Un programme, « Knowledge Seeker », aide beaucoup pour répondre à cette question. Il effectue automatiquement une analyse de la base de données et indique les éléments statistiques dont il y trouve. Nul besoin d'être un statisticien pour l'utiliser. Vous devez par contre décider si ses découvertes ont une certaine valeur archivistique. Le prix d'achat du programme est d'environ 600 \$.

Un autre problème, abondamment discuté, concerne le manque de normes relatives aux textes informatiques. La prolifération des formats, le manque d'organisation des documents informatiques, la difficulté à établir la distinction entre « routine » et « important » sont tous des aspects dont on s'est souvent plaint. Entre-temps, la plupart des organismes au Canada utilisent des ordinateurs pour effectuer leurs travaux. L'expérience acquise avec les fichiers du Bureau des négociations commerciales prouve que les documents informatiques peuvent être acquis et mis à la disposition des chercheurs. Un logiciel comme « Magellan » (180 \$) peut maîtriser les formats. Etant donné qu'il effectue la recherche du texte intégral (recherche de tout le contenu des documents), l'absence d'un numéro de dossier ou d'un autre système de classement ne pose pas de problème. Il est donc moins coûteux de prendre le fonds complet plutôt que d'essayer de séparer et transférer les documents importants.

Un troisième obstacle lié aux archives informatiques concerne la conservation. L'approvisionnement d'une bande magnétique, qui nécessite l'installation d'un système de climatisation, le rembobinage et la recopie. Le disque optique constitue une solution de rechange. Les mémoires de disque optique existent pour les micro-ordinateurs. Les disques eux-mêmes peuvent emmagasiner d'importantes quantités de données. Ils n'ont besoin d'aucun environnement ou traitement spécial, et ont une durée de vie d'environ cinquante ans (beaucoup plus que la durée de vie des machines dans lesquelles ils sont utilisés). Soit, cela est inférieur à la durée du papier, mais la durée de vie plus courte est en partie compensée par des coûts moins élevés d'entreposage. Un avertissement au sujet du disque optique s'impose : assurez-vous que celui que vous choisissez est conforme à une norme internationale, parce qu'il existe encore un grand nombre de formats brevetés, et qu'il vaut mieux ne pas les utiliser dans les archives informatiques.

Divers

Jusqu'à maintenant, j'ai parlé de certains obstacles rencontrés lors de la création d'un système efficace d'archives informatiques, et de la façon dont l'évolution de la technique des micro-ordinateurs les a diminués ou éliminés. Dans toute tentative de ce genre, les personnes en cause constituent le facteur le plus important. En examinant les quinze ans d'expérience des AN dans le domaine des archives informatiques, j'ai fait appel à des personnes détachées d'une expertise technique. Comment combiner cette expertise aux archives est un point à débattre. L'approche américaine (National Archives and Records Administration) a été de créer une unité spéciale ayant la pleine responsabilité des décisions techniques et archivistiques. Au sein de la Division des archives gouvernementales, une entité a été créée pour examiner les problèmes techniques, tandis que les archivistes s'occupaient des questions relatives aux archives. Les deux approches ont des avantages, mais la séparation des questions techniques et archivistiques permet plus facilement d'établir l'expertise technique nécessaire. Les bons techniciens sont très demandés de nos jours.

Qui sont les « bons techniciens » ? Pour moi, trois éléments s'avèrent importants. Je préfère quelqu'un qui aime les jeux d'aventures électroniques. La réalité dans le domaine de l'acquisition d'archives informatiques est la suivante : plus de la moitié des données ne sont pas ce qu'elles devraient être. La plus grande part du travail technique consiste donc à jouer un rôle d'enquêteur afin de déterminer ce que vous avez vraiment. Cela ressemble beaucoup à la résolution de casse-têtes dans un jeu d'aventures. Ensuite, je recherche une personne qui connaît un grand nombre de systèmes informatiques. L'administration fédérale a acheté au moins un ordinateur de chaque type vendu en Amérique du Nord. Connaitre IBM, Digital ou Sperry de manière approfondie est moins utile pour les archives que d'en savoir un peu au sujet des trois (et de plusieurs autres). Enfin, j'aimerais avoir quelques indices qui me prouveraient que la personne est en partie autodidacte. Tous les établissements d'archives doivent se tenir à jour en ce qui concerne les techniques utilisées par leurs donateurs et suivre l'évolution informatique en général. Les cours en informatique coûtent cher, et je n'en connais aucun qui réponde efficacement aux exigences quotidiennes des archives informatiques. Dans tous les cas, ce sont toujours les personnes qui font la réussite d'un programme.